

全国消費実態調査の 匿名データ4年分の統合メタデータ作成

高橋 行雄 BioStat研究所(株)

全消4年次分の擬似マイクロデータ

- ◆ 全国消費実態調査(全消)の1989, 1994, 1999 および 2004年の匿名データから作成した統計表を用い擬似マイクロデータを作成.
- ◆ この擬似マイクロデータを2019年9月5~6日に開催されたSASユーザー総会(@国際医療福祉大学・東京赤坂キャンパス)で実施した「Let'sデータ分析コンテスト」に供した.

匿名データでの実証研究

- ◆ 擬似マイクロデータを使って解析を行なうことは、匿名データを用いた本格的なデータ分析を行なうためのトレーニングとして最適.
- ◆ コンテストの最優秀賞の3組に翌年以降のSASユーザー総会で匿名データを使用した研究発表をお願いしてきた.
- ◆ 全消の匿名データの使用に際して、各種のサポートを行ってきた.

SASデータセット作成プログラム

- ◆ 統計センターから提供されていた教育用擬似マイクロデータを使ったコンテスト(2013～2016年)でのサポート.
- ◆ 固定長カンマ区切りの教育用擬似マイクロデータからSASデータセットを作成し、符号表から変数ラベルと変数フォーマットを自動作成するSASプログラムを提供.

全消・匿名データについても

- ◆ 全消の匿名データ(1994, 1999, 2004年度)を使った実証研究の便宜を図るため、匿名データからSASのリレーショナルデータベースを完全自動構築するSASプログラムの提供.
- ◆ 4年次分の擬似マイクロデータをコンテストに提供した経験から、更なる支援が必要と痛感.

2019年11月15日 高橋行雄


5

全消の匿名データのメタデータ

■ 全国消費実態調査 (最終更新日:2009/4/1)

年次	調査の概要	データレイアウト・符号表	レコード数	地域区分	【参考】 項目名一覧
平成元年 (1989年)	(報告書参照)	普通世帯 Excel:1.1MB	約4.5万レコード		CSV:87KB
		単身世帯 Excel:1.1MB	約0.3万レコード		CSV:87KB

:

平成16年 (2004年)	調査ページ (総務省統計局) 	二人以上の世帯 Excel:1.4MB	約4.4万レコード		CSV:82KB
		単身世帯 Excel:1.4MB	約0.4万レコード		CSV:82KB

2019年11月15日 高橋行雄

6

<https://www.nstac.go.jp/services/anonymity-zensho.html>

「データレイアウト・符号表」

- ◆ データレイアウト,2004年(二人以上の世帯)
 - 100 byteごとに233セット繰り返され, 78ページ
- ◆ 符号表
 - 項目名, 項目番号, 符号, 符号内容など
 - 項目番号は, ところどころ飛んでいて50ページ
- ◆ 参考・項目名一覧
 - 変数名・項目名(1786変数)
 - 75変数/ページとして24ページ

異なる変数番号(付表A, F)

- ◆ 擬似マイクロデータ4年分
 - 収支変数(Y001:年間収入~Y203:繰越金)
- ◆ 匿名:2004年
 - (V0399~V0622)の中に飛び飛びに存在
- ◆ 匿名:1994年
 - (U0405~U0619)の中に飛び飛びに存在

付表 E・収支に関連する年次別の変数名リスト (2/4)

S1989 変数名	T1994 変数名	U1999 変数名	V2004 変数名	Y共通 変数名	項目の レベル	V or S項目名 Y項目名	
S0789	T0555	U0458	V0454	Y039	2	実支出	
S0790	T0556	U0459	V0455	Y040	3	消費支出	
S0791	T0557	U0460	V0456	Y041	4	食料	

世帯員の繰返し

- ◆ 匿名データでは、7人の世帯員に関する項目が列方向に展開されているので、行方向に展開する。（表5）

世帯員 V0039～V0136 を 14変数区切りで行方向に展開									
	FT	V0003	V0004	データ1	データ2	データ3	データ4	...	データ14
世帯員	FT	Year V0003	No V0004	V0039 世帯員 番号	V0040 続柄	V0041 性別	V0042 年齢 5階級	...	V0052 各種学校 ・塾等
1	F	2004	1	V0039: 1	V0040	V0041	V0042	...	V0052
2	F	2004	1	V0053: 2	V0054	V0055	V0056	...	V0066
:	:	:	:	:					
7	F	2004	1	V0123: 7	V0124	V0125	V0126	...	V0136

2019年11月15日 高橋行雄

11

微妙に異なる符号内容の壁

- ◆ 年度によって符号内容の表現が異なる
 - 各年度の符号表をデータベース化
 - 世帯区分: 1989年:S0014, 1994年:T0015
1999年:U0015, 2004年:V0015
 - 共通変数: Y02 として符号内容を並び替える

VUTS項目名	Y変数名	Y符号	1989	1994	1999	2004
世帯区分	Y02	1	勤労者世帯	勤労者世帯	勤労	勤労
世帯区分	Y02	2	一般世帯	無職を除く勤労者以外の世帯	勤労以外	勤労以外
世帯区分	Y02	3	無職世帯	無職世帯	無職	無職

2019年11月15日 高橋行雄

12

職業符号の場合

変数名	項目名	符号WXY	1989	1994	1999	2004
V48	職業符号	0	その他	その他	自由業者	
V48	職業符号	1	△1 常用労務作業者	常用労務作業者	常用労務作業者	常用労務作業者
V48	職業符号	2	△2 臨時及び日々雇 労務作業者	臨時及び日々雇 労務作業者	臨時及び日々雇 労務作業者	臨時及び日々雇 労務作業者
V48	職業符号	3	△3 民間職員	民間職員	民間職員	民間職員
V48	職業符号	4	△4 官公職員	官公職員	官公職員1	官公職員1
V48	職業符号	5	△5 商人及び職人	商人及び職人	官公職員2	官公職員2
V48	職業符号	6	△6 個人経営者	個人経営者	商人及び職人	商人及び職人
V48	職業符号	7	△7 農林漁業従業者	農林漁業従業者	個人経営者	個人経営者
V48	職業符号	8	△8 法人経営者	法人経営者	農林漁業従業者	農林漁業従業者
V48	職業符号	9	△9 自由業者	自由業者	法人経営者	法人経営者
V48	職業符号	W(12)			その他	
V48	職業符号	X(10)	無職	無職	無職	
V48	職業符号	Y(11)	家族従業者	家族従業者	家族従業者	
V48	職業符号		10			自由業者
V48	職業符号		11			その他
V48	職業符号		12			無職
V48	職業符号		13			家族従業者

2019年11月15日 高橋行雄

13

世帯7分類(表6)

◆ 世帯員データから新たに生成

分類内容	FT	1989 年次	1994 年次	1999 年次	2004 年次	計
1_単身	T	3,236	3,813	3,982	3,936	14,967
2_夫婦のみ	F	8,784	10,510	12,422	13,447	45,169
	T	6	0	0	0	6
3_二世代	F	22,988	22,629	21,379	19,982	86,978
4_二世代(世帯主と子)	F	1,302	1,676	1,892	2,260	7,131
	T	1	0	0	0	1
5_二世代(世帯主または 夫婦とひとり親)	F	1,379	1,534	1,758	1,983	6,654
6_三世代	F	8,751	7,295	6,099	5,139	27,284
7_その他	F	1,333	1,043	990	1,050	4,416
計		47,780	48,500	48,522	47,797	192,606

2019年11月15日 高橋行雄

14

続柄の再分類

元の符号	変換後 符号	符号内容	10進数	Z変数 0 or 1	続柄
1	1	世帯主	0	Z1	直系親族
2	2	世帯主の配偶者	1	Z2	
4	3	子(未婚)	2	Z3	
4	4	子(夫婦であるもの)			
8	8	孫(未婚)	4	Z4	
9	9	孫(夫婦であるもの)			
5	5	父母	8	Z5	
0	0	祖父母	16	Z6	
6	6	兄弟(未婚)	32	Z7	傍系親族
7	7	兄弟(夫婦であるもの)			
X	10	他の親族			
Y	11	使用人・同居人			非親族

2進数を10進数へ変換 = $Z2 + 2 \times Z3 + 4 \times Z4 + 8 \times Z5 + 16 \times Z6 + 32 \times Z7$

続柄の再分類の方法

- ◆ 年度別・TF別・世帯No別に変換後の符号(0~11)についての集計表ファイルを作成。
 - 変数名として F0, F1, ..., F11
 - 世帯毎に変数(F0~F11)の人数をカウント
 - 変数(F0~F11)を用いて、二値(0, 1)の変数(Z1~Z7)を計算
 - 二値変数(Z1~Z7)から10進数に変換
- ◆ 10進数別・Z変数別・年度別クロス表(付表D)を作成し、世帯分類を手作業で与える。

付表D 詳細な世帯分類表に準拠した年次毎の世帯数															
10進数 d	Z1 世帯主	Z2 配偶者	Z3 子	Z4 孫	Z5 父母	Z6 人数	Z7 祖父母	Z7 兄弟など							
									1989	1994	1999	2004	計	世帯分類	
0	1	0	0	0	0	0	0	0	3,236	3,813	3,982	3,936	14,967	単身世帯	
1	1	1	0	0	0	0	0	0	8,790	10,510	12,422	13,447	45,169	夫婦のみ	
2	1	0	1	0	0	0	0	0	1,303	1,676	1,892	2,260	7,131	2世代(世帯主と子)	
3	1	1	1	0	0	0	0	0	22,697	22,315	21,025	19,559	85,596	2世代	
4	1	0	0	1	0	0	0	0	29	26	30	17	102	三世代	
5	1	1	0	1	0	0	0	0	44	46	41	59	190	三世代	
6	1	0	1	1	0	0	0	0	197	168	146	139	650	三世代	
7	1	1	1	1	0	0	0	0	1,582	1,006	857	660	4,105	三世代	
8	1	0	0	0	1	1	0	0	379	446	543	616	1,984	2世代(世帯主 または夫婦とひとり親)	
8	1	0	0	0	1	2	0	0	99	113	144	163	519	2世代	
9	1	1	0	0	1	1	0	0	1,000	1,088	1,215	1,367	4,670	2世代(世帯主 または夫婦とひとり親)	
9	1	1	0	0	1	2	0	0	192	201	209	259	861	2世代	
9	1	1	0	0	1	3	0	0	0	0	1	1	2	2世代	
56	1	0	0	0	1		1	1	13	11	15	6	45	他+三世代	
57	1	1	0	0	1		1	1	1	2	1	1	5	他+三世代	
58	1	0	1	0	1		1	1	1	1	0	0	2	他+4世代	
59	1	1	1	0	1		1	1	4	1	4	2	11	他+4世代	
計									47,780	48,500	48,522	47,797	192,599		

統合したメタデータ

- ◆ 年度毎の匿名データからSASのリレーショナルデータベースを作成するSASプログラムの提供で十分ではないのか？
- ◆ 単年次の匿名データを対象とする場合には、有益ではあるが、複数年次の匿名データを使った実証研究を行う場合には、年次毎の異なる変数名の統合が不可欠。
- ◆ さらに、年次毎に微妙に異なる符号内容にも注意を払う必要もある。

どのような情報の提供が有益か

- ◆ 匿名データを(世帯, 世帯員, 収支)などに区分した場合の変数(番号)名の範囲
- ◆ (世帯, 世帯員, 収支)について年次毎の変数(番号)名の対応表
- ◆ 年度別・変数別・符号別・符号内容
 - このデータから, 変数別・符号別・年度別の符号内容の比較が容易になる

年度による異なる符号

- ◆ 調査年度毎に同じ符号だが, 内容が異なることは, 年とともに生活様式が異なることを反映している.
- ◆ 「住居の建て方」
 - 1989年, 1994年には, 「5:共同住宅6階建以上」であった.
 - 1999年, 2004年には, 「5:共同住宅は6~10階建」と「6:11階建以上」に変更.
 - その結果, 「その他」が(6, 7)と別符号.

符号の摺り合わせの必要性

- ◆ 擬似マイクロデータで「住居の建て方」
 - 「1:一戸建て」, 「2:一戸建て以外」と単純な符号とした.
 - 匿名データを使う場合は, 1989年と1994年の「6:その他」を「7:その他」に符号を変えることにより, 正確な集計が可能となる.
- ◆ このような符号の変換を正確に一括して行うための方法論(実践方法)が必要である.

列と行を入れ替えて一括変換

変換対象 世帯員ファイル						
世帯連番	年次	No	FT	世帯員 No	V47	V48
1	1989	1	F	1	1	0
2	1989	1	F	2	2	10
:						
:						

- ◆ 列方向の元符号を行方向に並べ変える.
- ◆ マッチマージで変換符号を付与する.

変数の積み上げ 変換対象 世帯員ファイル									
世帯連番	年次	No	FT	世帯員 No	変数名	元符号	符号内容	変換符号	
1	1989	1	F	1	V47	1	農業・林業・漁業	1	
1	1989	1	F	1	V48	0	その他	11	
1	1989	1	F	2	V47	2	鉱業	4	
1	1989	1	F	2	V48	10	無職	12	
:									
:									

- ◆ 行方向の変換符号を列方向に並べ変える.

匿名データの利用の活性化

- ◆ 複数年次の全国消費実態調査の匿名データから統計表を作成し、擬似マイクロデータを作成した。
- ◆ その経験から、匿名データの利用の活性化を図るためには、統合するために必要なメタデータの提供が不可欠と思われた。
- ◆ また、年度毎に異なる変数(番号)名の共通化の方法論、符号内容が異なる場合に共通符号を付与する方法論も必要と思われた。

ご清聴ありがとうございました